A New Approach in Fitting Linear Regression Models with the Aim of Improving Accuracy and Power

Soheil Sadeghi\textsuperscript{1*}, Hashem Mahlooji\textsuperscript{2}

Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran
\textsuperscript{1}s_sadeghi@ie.sharif.edu, \textsuperscript{2}mahlooji@sharif.edu

ABSTRACT

The main contribution of this work lies in challenging the common practice of inferential statistics in the realm of simple linear regression for attaining a higher degree of accuracy when multiple observations are available, at least, at one level of the regressor variable. We derive sufficient conditions under which one can improve the accuracy of the interval estimations at quite affordable extra computational cost. Two algorithms and a numerical example will be presented to fully explain how our approach works and to compare the results of our approach versus the results obtained from three of the well known statistical software systems.

Keywords: Simple linear regression, Multiple observations, Weighted least squares, Accuracy, Power

1. INTRODUCTION

Linear regression models are among the most popular statistical tools which have been successfully applied to a wide spectrum of problems. To apply regression models, the common practice is to collect data and minimize the sum of squared deviations between the observed data and the values coming from the underlying relation, Montgomery et al. (2001) and Neter et al. (1996). Once the regression model is developed, various tests of hypothesis as well as confidence and prediction intervals can be presented.

Here, we start with the simple linear regression model shown in (1) which we label as the Original Model throughout the article:

\[ Y = X \cdot \beta + \epsilon \]  

(1)

where

\[ Y = \begin{pmatrix} y_{11} \ldots y_{m_{1}1} \\ \vdots \\ y_{1c} \ldots y_{m_{c}c} \end{pmatrix}^{T} \]

\[ X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m_{1}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1c} & \cdots & x_{1m_{c}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm_{1}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{nc} & \cdots & x_{nm_{c}} \end{pmatrix} \]

\[ \beta = \begin{pmatrix} \beta_{0} \\ \beta_{1} \\ \vdots \\ \beta_{m_{1} - 1} \\ \beta_{m_{1}} \\ \vdots \\ \beta_{m_{c} - 1} \end{pmatrix} \]

\[ \epsilon = \begin{pmatrix} \epsilon_{1} \\ \epsilon_{2} \\ \vdots \\ \epsilon_{n} \end{pmatrix} \]

\[ \epsilon \sim N(0, \sigma^{2} I) \]

\* Corresponding Author

ISSN: 1735-8272, Copyright © 2010 JISE . All rights reserved.
\[
X = \left( \begin{array}{cccc}
1 & \cdots & 1 & 1 \cdots 1 \\
\frac{x_{11}}{m_1} & \cdots & \frac{x_{1c}}{m_c} \\
\frac{x_{21}}{m_1} & \cdots & \frac{x_{2c}}{m_c} \\
\vdots & \ddots & \vdots \\
\frac{x_{c1}}{m_1} & \cdots & \frac{x_{cc}}{m_c}
\end{array} \right)^T \\
\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \\
\epsilon = \left( \begin{array}{ccc}
\frac{\epsilon_{11}}{m_1} & \cdots & \frac{\epsilon_{11}}{m_1} \\
\frac{\epsilon_{12}}{m_2} & \cdots & \frac{\epsilon_{12}}{m_2} \\
\vdots & \ddots & \vdots \\
\frac{\epsilon_{1c}}{m_c} & \cdots & \frac{\epsilon_{1c}}{m_c}
\end{array} \right)^T
\]

In this model there are \( m_j \), \( j = 1,2,3,\ldots,c \) observations from the response variable, \( Y \), at each of the \( c \) levels of the regressor variable, \( X \), such that there are \( n \) pairs of data altogether (see Figure 1). For the sake of our discussion it suffices for just one \( m_j \) to be strictly larger than 1.

This means that \( m_j \in \mathbb{N} \), \( \sum_{j=1}^{c} m_j = n \)

![Figure 1 The layout of data in the Original Model](image)

The commonly adopted assumptions in regression analysis include:

1. \( E(\epsilon_i) = 0 \quad ; \quad i = 1,2,\ldots,n \) (2)
2. \( \text{Var}(\epsilon_i) = \sigma^2 \quad (\text{unknown}) \quad ; \quad i = 1,2,\ldots,n \) (3)
3. \( \text{Cov}(\epsilon_i, \epsilon_{i'}) = 0 \quad ; \quad i, i' = 1,2,\ldots,n \quad i \neq i' \) (4)

Assuming that \( \epsilon_i \)'s behave as normal random variables, appropriate confidence and prediction intervals can be presented. Since it is possible that the unknown variances in (3) not to be fixed for all \( c \)'s, one can adopt the weighted least squares to treat such a case.

The commonly adopted approach toward statistical inference in the regression problem as outlined above, often fails to exploit all the potential that the collected data has to offer in order to arrive at the best possible results. In this work we analytically establish the conditions under which such potential is practically wasted while by bearing a very affordable cost one could have achieved more accurate results. In the quest for identifying the optimal (regression) model, i.e., the model that better exploits the potential of the collected data, we present two algorithms and by borrowing a typical problem from literature we extensively demonstrate how our proposed approach works and what it can achieve when compared with the approach which is commonly practiced.

This paper is organized in the following manner: section 2 investigates a model in which the mean of observations at just one level of the independent variable takes the place of the observations.
themselves. We examine to see if relations (2), (3), and (4) are still true for such a case and then arrive at the regression model. Section 3 presents a similar discussion when the means of observations at more than one level of the independent variable are considered. Section 4 presents an algorithm for finding a model with the smallest mean squared error ($MSE$). Section 5 is devoted to generating a model with the largest accuracy. Section 6 includes a numerical example. Section 7 concludes the discussion and presents some ideas for further investigation.

2. REGRESSION ON THE MEAN OF OBSERVATIONS AT A SINGLE LEVEL

In this section we assume that at a single level of the regressor variable $X$ we have available $m_j > 1$, $j = 1, 2, ..., c$ observations on $Y$. Since it is irrelevant at which level of $X$ these multiple observations are collected, here we assume that the $m_j$ observations correspond to the level $x_j$ where $j = 1$ and we intend to employ the mean of these observations in fitting the regression model. As such, we are dealing with the following model which we label as Model $j$ where $j = 1$ (unless specified otherwise).

$$Y_1 = X_1\beta + \varepsilon_1$$

(5)

in which

$$Y_1 = \left( \frac{\sum_{k=1}^{m_1} y_{k1}}{m_1} \quad \frac{y_{12} \cdots y_{m_2}}{m_2} \cdots \quad \frac{y_{1c} \cdots y_{mc}}{m_c} \right)^T$$

is an $(n - m_1 + 1) \times 1$ matrix,

$$X_1 = \begin{pmatrix} x_1 & 1 \ldots 1 & 1 \ldots 1 \\ 0 & x_2 \cdots x_2 & x_c \cdots x_c \end{pmatrix}^T$$

is an $(n - m_1 + 1) \times 2$ matrix,

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

and finally,

$$\varepsilon_1 = \left( \frac{\sum_{k=1}^{m_1} \varepsilon_{k1}}{m_1} \quad \frac{\varepsilon_{12} \cdots \varepsilon_{m_2}}{m_2} \cdots \quad \frac{\varepsilon_{1c} \cdots \varepsilon_{mc}}{m_c} \right)^T$$

is an $(n - m_1 + 1) \times 1$ matrix.

To arrive at the regression model in (5) we first present the following definitions:

$S_1$: represents an $(n - m_1 + 1) \times n$ matrix of the form

$$S_1 = \begin{pmatrix} \frac{m_1}{1} & 0 & \ldots & 0 \\ 0 & \frac{m_1}{0} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \frac{n-m_1}{1} \end{pmatrix}$$
where in the first row the first $m_1$ entries are 1 and the remaining entries are 0. In the lower right side of $S_1$ there is an $(n - m_1) \times (n - m_1)$ unit sub-matrix and all other entries of $S_1$ are 0. $W_1$: stands for an $(n - m_1 + 1) \times (n - m_1 + 1)$ matrix of the form

$$W_1 = \begin{pmatrix}
\frac{1}{m_1} & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}$$

where each of its off diagonal entries is zero, and all diagonal entries are $1$ except for the entry at northwest corner which is $1/m_1$.

We have defined $S_1$ and $W_1$ in a way such that by multiplying both sides of (1) by $W_1 S_1$ one can arrive at $\mathbf{1}$.

### 2.1. Parameter Estimation

We first examine to see if (2), (3), and (4) hold in $\mathbf{1}$. As can be seen below, (2) and (4) hold true while this is not the case for (3). To get around this problem we resort to weighted least squares to fit our model.

$$E(\varepsilon_1) = \frac{1}{m_1} \sum_{k=1}^{m_1} E(\varepsilon_{k1}) = 0 \quad \text{where} \quad \varepsilon_1 = \frac{1}{m_1} \sum_{k=1}^{m_1} \varepsilon_{k1}$$

$$E(\varepsilon_i) = 0 \quad ; \quad i = 2, \ldots, n - m_1 + 1$$

$$Var(\varepsilon_i) = \frac{1}{m_1} \sum_{k=1}^{m_1} Var(\varepsilon_{k1}) = \frac{\sigma^2}{m_1}$$

$$Var(\varepsilon_i) = \sigma^2 \quad ; \quad i = 2, \ldots, n - m_1 + 1$$

$$Cov(\varepsilon_i, \varepsilon_i') = \frac{1}{m_1} \sum_{k=1}^{m_1} Cov(\varepsilon_{k1}, \varepsilon_{i'1}) = 0 \quad ; \quad i' = 2, \ldots, n - m_1 + 1$$

$$Cov(\varepsilon_i, \varepsilon_i') = 0 \quad ; \quad i, i' = 2, \ldots, n - m_1 + 1 \quad i \neq i'$$

Therefore

$$E(\varepsilon_1) = [0 \quad 0 \quad \ldots \quad 0]^T \quad \text{and} \quad Var(\varepsilon_1) = \sigma^2 \begin{pmatrix}
\frac{1}{m_1} & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix} = \sigma^2 W_1$$

It is clear that the weighted least squares (WLS) estimators of the parameters in Model 1 can be written as Neter et al. (1996).

$$\mathbf{b}_1 = (X_1^T W_1^{-1} X_1)^{-1} X_1^T W_1^{-1} Y_1$$

(6)

Now, we establish the fact that $\mathbf{b}_1$ is exactly the same as $\mathbf{b} = [X^T X]^{-1} X^T Y$ which represents the least squares (LS) estimators of the Original Model. To show this we substitute $X_1$ and $Y_1$ in (6) by $W_1, S_1, X$ and $W_1, S_1, Y$ respectively to obtain

$$\mathbf{b}_1 = \left[(W_1 S_1 X)^T W_1^{-1} (W_1 S_1 X)\right]^{-1} (W_1 S_1 X)^T W_1^{-1} (W_1 S_1 Y)$$

$$= [X^T S_1^T W_1^{-1} W_1^T S_1 X]^{-1} X^T S_1^T W_1^{-1} W_1^T S_1 Y$$
noting that $W_1^T = W_1$ and $W_1^{-1}W_1 = I$, we will have $S_1^T W_1^{-1}W_1S_1 = S_1^T W_1S_1$

$$
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
\frac{1}{m_1} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{m_1} & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
1 & \ldots & 1 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 1
\end{pmatrix}
$$

$$
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
\frac{1}{m_1} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{m_1} & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
1 & \ldots & 1 & 0 & \ldots & 0 \\
0 & \ldots & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 1
\end{pmatrix}
$$

$$
\begin{pmatrix}
\frac{1}{m_1} & \ldots & \frac{1}{m_1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{m_1} & \ldots & \frac{1}{m_1} \\
0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 1
\end{pmatrix}
$$

Designating this matrix by $A_1$, we have

$$
X^T A_1 = \begin{pmatrix}
1 & \ldots & 1 \\
\frac{1}{x_1} & \ldots & \frac{1}{x_1} \\
\frac{1}{x_2} & \ldots & \frac{1}{x_2} \\
\frac{1}{x_{e_1}} & \ldots & \frac{1}{x_{e_1}}
\end{pmatrix}
\begin{pmatrix}
\frac{1}{m_1} & \ldots & \frac{1}{m_1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{m_1} & \ldots & \frac{1}{m_1} \\
0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 1
\end{pmatrix} = X^T
$$

therefore, having in mind that $S_1^T W_1S_1 = A_1$ and $X^T A_1 = X^T$, one can write

$$
b_1 = \left[X^T S_1^T W_1 S_1 X \right]^{-1} X^T S_1^T W_1 S_1 Y = \left[X^T A_1 X \right]^{-1} X^T A_1 Y =
\left[X^T X \right]^{-1} X^T Y = b
$$

As can be seen, the LS estimators for Model 5 are exactly the same as those obtained for the Original Model. The variance of the coefficients in the Original Model is

$$
Var(b) = \sigma^2 (X^T X)^{-1}
$$

and since $b_1$ in Model 1 is the same as $b$ (in the Original Model), then obviously $Var(b_1) = Var(b)$.

2.2. The Sum of Squares

2.2.1. The Sum of Squared Error (SSE)

So far we have shown that the estimated regression line in Model 1 coincides with the one in the Original Model. By examining Figure 2
It is evident that the sum of squared errors (SSE) differ for the two models only at level $x_1$. In fact, SSE’s for the two models have identical terms at all the other levels of the regressor variable. Let $SSE_0$ and $SSE_1$ designate the sum of squared errors in the Original Model and the proposed model (Model 1) respectively. Thus we can write

$$SSE_0 = \sum_{j=1}^{c} \sum_{k=1}^{m_j} (y_{kj} - \hat{y}_j)^2 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2 + \sum_{j=2}^{c} \sum_{k=1}^{m_j} (y_{kj} - \hat{y}_j)^2$$

$$SSE_1 = m_1. (\bar{y}_1 - \hat{y}_1)^2 + \sum_{j=2}^{c} \sum_{k=1}^{m_j} (y_{kj} - \hat{y}_j)^2$$

Where $\hat{y}_j$ stands for the expected value of the dependent variable at level $j$ of the regressor variable. To set up a relation between $SSE_0$ and $SSE_1$, we can write

$$SSE_0 - SSE_1 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2 - m_1 (\bar{y}_1 - \bar{y}_1)^2$$

$$= \sum_{k=1}^{m_1} (y_{k1}^2 + \bar{y}_1^2 - 2y_{k1}\bar{y}_1) - m_1 (\bar{y}_1^2 + \bar{y}_1^2 - 2\bar{y}_1\bar{y}_1)$$

$$= \sum_{k=1}^{m_1} y_{k1}^2 - m_1\bar{y}_1^2 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2$$

this means that $SSE_0 = SSE_1 + S\gamma_1$, where $S\gamma_1 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2$.

In other words, when the mean of $m_1$ observations is used instead of the $m_1$ observations themselves, SSE will be reduced by as much as $S\gamma_1 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2$.

### 2.2.2. The Total Sum of Squares (SST)

To compare SST for the Original Model and Model 5 we first calculate the mean of the observed values of the response variable for Model 5 as

$$\bar{y}^{(5)} = \frac{1}{m_1+n-m_1} (m_1\bar{y}_1 + \sum_{i=m_1+1}^{n} y_i) = \frac{1}{n} (\sum_{k=1}^{m_1} y_{k1} + \sum_{j=2}^{c} \sum_{k=1}^{m_j} y_{kj})$$

$$= \frac{1}{n} (\sum_{j=1}^{c} \sum_{k=1}^{m_j} y_{kj}) = \bar{y}$$
As expected, the mean of the observed values of \( Y \) in both models are identical. Now, according to the situations displayed in Figure 3, \( SST \) for the Original Model, \( SST_0 \), and the proposed model, \( SST_1 \), can be written as

\[
SST_0 = \sum_{j=1}^c \sum_{k=1}^{m_j} (y_{kj} - \bar{y})^2 = \sum_{k=1}^{m_1} (y_{1j} - \bar{y})^2 + \sum_{j=2}^c \sum_{k=1}^{m_j} (y_{kj} - \bar{y})^2
\]

\[
SST_1 = m_1 (\bar{y}_1 - \bar{y})^2 + \sum_{j=2}^c \sum_{k=1}^{m_j} (y_{kj} - \bar{y})^2.
\]

These \( SST \) differ by as much as

\[
SST_0 - SST_1 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y})^2 - m_1 (\bar{y}_1 - \bar{y})^2
\]

\[
= \sum_{k=1}^{m_1} (y_{k1}^2 + \bar{y}^2 - 2y_{k1}\bar{y}) - m_1 (\bar{y}_1^2 + \bar{y}^2 - 2\bar{y}_1\bar{y})
\]

\[
= \sum_{k=1}^{m_1} y_{k1}^2 - m_1 \bar{y}_1^2 = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2
\]

which means that

\[
SST_0 = SST_1 + S_{Y_1},
\]

where \( S_{Y_1} = \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2 \), as before.

The last equation above reveals that by regressing on the mean of the observed values of \( Y \) at level \( x_1 \), \( SST \) tends to decrease as much as \( S_{Y_1} \).

2.2.3. The Sum of Squares due to Regression (SSR)

Since \( SST = SSE + SSR \) holds so does the relation \( SSR_1 = SSR_0 \), which means that \( SSR \) is the same for both models.

2.3. The Mean Squared Error (MSE)

To compare the MSE’s of the Original Model and Model 1, i.e., \( MSE_0 \) and \( MSE_1 \), we can write

\[
SSE_0 = SSE_1 + S_{Y_1}
\]

or
\[(n - 2)MSE_0 = (n - m_1 - 1)MSE_1 + S_{Y_1}\]

or finally
\[MSE_1 = \frac{1}{(n-m_1-1)}[(n - 2)MSE_0 - S_{Y_1}]\]

This means that for the proposed model to reduce the \(MSE\), it is sufficient to have
\[\frac{1}{(n-m_1-1)}[(n - 2)MSE_0 - S_{Y_1}] < MSE_0\]

or
\[MSE_0 < \frac{1}{m_1-1}S_{Y_1}.\]

Or, finally, we have
\[MSE_0 < S_1^2\] \hspace{1cm} (7)

where
\[S_1^2 = \frac{1}{m_1-1} \sum_{k=1}^{m_1} (y_{k1} - \bar{y}_1)^2.\]

It is interesting to note that both \(MSE_0\) and \(MSE_1\) are unbiased estimators of \(\sigma^2\) and in case the relation (7) holds, the \(MSE\) in the proposed model will be smaller than that in the \textit{Original Model}. Besides, as shown earlier, the variances of the coefficients in both models are identical and equal to

\[Var(b) = \sigma^2(X^T.X)^{-1}\]

Employing the unbiased estimators of \(\sigma^2\) here, leads to the following point estimators for variances of the coefficients:

\textit{Original Model:} \hspace{1cm} \[S_0^2(b) = MSE_0(X^T.X)^{-1}\]

\textit{Model 1:} \hspace{1cm} \[S_1^2(b) = MSE_1(X^T.X)^{-1}\]

It is clear that forcing \(MSE_0 < S_1^2\) ensures us that the estimators of variance of the coefficients in the proposed model tend to decrease in comparison to the corresponding estimates from the \textit{Original Model}.

\textbf{2.4. Coefficient of Determination} \(R^2\)

As far as the coefficient of determination is concerned, one can write
\[R_0^2 = \frac{SSR_0}{SST_0} \quad \text{and} \quad R_1^2 = \frac{SSR_1}{SST_1}\]

for the original and the proposed models. Now, the following relation shows that employing the proposed model always tends to increase \(R^2\).
\[ R_1^2 = \frac{SSR_1}{SST_1} = \frac{SSR_0}{SST_0 = SY_1} > \frac{SSR_0}{SST_0} = R_0^2 \]

2.5. Power of the Tests and the Length of the Intervals

To judge the power of the tests as well as the length of the confidence and prediction intervals, in this section we assume that the error terms behave as normal random variables.

2.5.1. Judging the Parameter Estimators

The two-sided \( 100(1 - \alpha)\% \) confidence interval for parameter \( \beta_l, l = 1,2 \) for the proposed model can be written as

\[ b_l \pm t_{1 - \frac{\alpha}{2}, n - m_1 - 1} \cdot S(b_l), \ l = 0,1 \]

where \( S(b_l) \) is the square root of the estimated variance of the point estimator of \( \beta_l \).

The length of this interval reflects upon the interval’s accuracy and the power of the corresponding test in the sense that the shorter this interval, the higher the accuracy as well as the power would be. Employing the proposed model will decrease the degrees of freedom and in spite of the fact that \( MSE_0 < S_1^2 \), the value of \( MSE \) and consequently the values of \( S(b_l) \) tend to decrease. As such, for comparing the accuracy and the power in employing the two models it would be sufficient to compare the term

\[ t_{1 - \frac{\alpha}{2}, d.f.} \cdot \sqrt{MSE} \quad (8) \]

in the two models. Whenever this term for one of the models is smaller, that model provides a higher degree of accuracy and more power.

2.5.2. Inferring on the Mean of the Response Variable at a Fixed Level of the Independent Variable

A two-sided \( 100(1 - \alpha)\% \) confidence interval for the expected value of the response variable at the level \( x_h \) assumes the following form

\[ \hat{y}_h \pm t_{1 - \frac{\alpha}{2}, n - m_1 - 1} \cdot S(\hat{y}_h) \quad (9) \]

where \( \hat{y}_h \) represents the value that the estimated line assumes when \( x = x_h \) and \( S^2(\hat{y}_h) \) is defined as

\[ S^2(\hat{y}_h) = MSE.\mathbf{X}_h^T(\mathbf{X}^T\mathbf{X})^{-1}.\mathbf{X}_h \]

in which \( \mathbf{X}_h = \begin{pmatrix} 1 \\ x_h \end{pmatrix} \).

A comparison of confidence intervals presented in (9) for the two regression models leads again to the comparison of (8) for these models.
2.5.3. Inferring on a New Observation of the Response Variable at a Specified Level of $X$

The prediction interval for a new observation of the response variable when $x$ is considered at level $x_h$ can be written as

$$\hat{y}_h + t_{1 - \frac{\alpha}{2}, n - m_1 - 1} \cdot S(pred)$$

where

$$S(pred) = \left(\text{MSE} \cdot \left[1 + X_h^T (X^T X)^{-1} X_h\right]\right)^{\frac{1}{2}}$$

Once again, comparing this interval for the two models leads to a comparison of (8) for them.

Our findings so far indicate that decreasing $MSE$ by itself is not a sufficient yardstick for attaining more accuracy and power and we showed instead that we should see to it that the following holds true:

$$t_{1 - \frac{\alpha}{2}, n - m_1 - 1} \cdot \sqrt{MSE_1} < t_{1 - \frac{\alpha}{2}, n - 2} \cdot \sqrt{MSE_0}$$

or

$$\frac{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}{t_{1 - \frac{\alpha}{2}, n - 2}} \sqrt{\frac{1}{(n - m_1 - 1)} [(n - 2)MSE_0 - S_{Y_1}]} < \sqrt{MSE_0}$$

or

$$\frac{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}{t_{1 - \frac{\alpha}{2}, n - 2}} \sqrt{\frac{1}{(n - m_1 - 1)} [(n - 2) - \frac{S_{Y_1}}{MSE_0}]} < 1$$

or

$$\sqrt{\frac{1}{(n - m_1 - 1)} [(n - 2) - \frac{S_{Y_1}}{MSE_0}]} < \frac{t_{1 - \frac{\alpha}{2}, n - 2}}{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}$$

or

$$(n - 2) - \frac{S_{Y_1}}{MSE_0} < (n - m_1 - 1) \left(\frac{t_{1 - \frac{\alpha}{2}, n - 2}}{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}\right)^2$$

or

$$(n - 2) - (n - m_1 - 1) \left(\frac{t_{1 - \frac{\alpha}{2}, n - 2}}{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}\right)^2 < \frac{S_{Y_1}}{MSE_0},$$

and, finally, we have

$$\left[1 - \frac{(n - m_1 - 1)}{(n - 2)} \left(\frac{t_{1 - \frac{\alpha}{2}, n - 2}}{t_{1 - \frac{\alpha}{2}, n - m_1 - 1}}\right)^2\right]SSE_0 < S_{Y_1}.$$  

(10)
This means that once (10) holds, the proposed approach generates more powerful tests and more accurate confidence and prediction intervals.

3. EMPLOYING OBSERVATIONS’ MEANS AT MORE THAN ONE LEVEL

3.1. Means at Two Levels

Let \( x_{j_1} \) and \( x_{j_2} \) represent the levels at which multiple measurements from the response variable are available. We intend to employ means of such observations in fitting the regression model. To achieve this purpose one can develop the intended model based on either the Original Model or Model 5. In fact, the model will look like the following which we label as Model \( j_1, j_2 \), \( j_1 = 1, j_2 = 2 \) meaning the model in which means are employed at levels \( x_{j_1} \) and \( x_{j_2} \) of the regressor variable.

\[
Y_{1,2} = X_{1,2} \beta + \epsilon_{1,2}
\]

in which \( Y_{1,2} \) and \( \epsilon_{1,2} \) represent \((n - m_1 - m_2 + 2) \times 1\) matrices, i.e.,

\[
Y_{1,2} = \begin{pmatrix} \bar{y}_1 & \bar{y}_2 & y_{13} & \cdots & y_{m_2} \\ m_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_c \\ \end{pmatrix}^T
\]

and

\[
\epsilon_{1,2} = \begin{pmatrix} \bar{\varepsilon}_1 & \bar{\varepsilon}_2 & \varepsilon_{13} & \cdots & \varepsilon_{m_2} \\ m_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_c \\ \end{pmatrix}^T
\]

\[
\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \end{pmatrix}
\]

and \( X_{1,2} \) represents the following \((n - m_1 - m_2 + 2) \times 2\) matrix

\[
X_{1,2} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 1/m_2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \\ \end{pmatrix}^T
\]

Here, \( \bar{y}_1, \bar{y}_2, \bar{\varepsilon}_1, \) and \( \bar{\varepsilon}_2 \) represent the means of the observed data as well as the means of error terms at levels \( x_1 \) and \( x_2 \) respectively.

To arrive at the model in (11), it suffices to multiply both sides of (5) by the matrix combination

\[
EF = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1/m_2 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \end{pmatrix} \cdot \begin{pmatrix} m_2 \\ 0 \\ \vdots \\ 0 \\ \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \end{pmatrix}
\]

where \( E \) and \( F \) are \((n - m_1 - m_2 + 2) \times (n - m_1 - m_2 + 2)\) and \((n - m_1 - m_2 + 2) \times (n - m_1 + 1)\) matrices. It is quite straightforward to show that relations (2) and (4) hold for the model in (11). One can resort to the weighted least squares to get around the nonhomogeneity of variances of the errors.

By defining the \((n - m_1 - m_2 + 2) \times (n - m_1 - m_2 + 2)\) matrix \( W_{1,2} \) as
the weighted least squares point estimators of the coefficients can be obtained. Specifically, in suffices for \( W_{1,2} \) to take the place of \( W_1 \) in (6) and \( X_1 \) and \( Y_1 \) be substituted by \( X_{1,2} \) and \( Y_{1,2} \) respectively. Here, again it will be a simple matter to show that the point estimators are exactly the same as the estimators in Model 1 and hence, in the Original Model.

Similar to what was shown, again it can be shown that

\[
SSE_{1,2} = SSE_1 - S_{Y_2} = SSE_0 - S_{Y_1} - S_{Y_2}
\]

and

\[
SST_{1,2} = SST_1 - S_{Y_2} = SST_0 - S_{Y_1} - S_{Y_2}
\]

and

\[
SSR_{1,2} = SSR_1 = SSR_0.
\]

As before, when \( MSE_1 < S_{2}^2 = \frac{1}{m_2-1} \sum_{k=1}^{m_2} (y_{k2} - \bar{y}_2)^2 \) holds, employing the model in (11) tends to decrease the \( MSE \) when compared to Model 1. Now we derive the conditions under which the model in (11) provides a smaller \( MSE \) compared to the Original Model.

In order for the inequality \( MSE_{1,2} < MSE_0 \) to be true, we should have

\[
\frac{n-2}{n-m_1-m_2} (MSE_0 - S_{Y_1} - S_{Y_2}) < MSE_0
\]

or

\[
MSE_0 < \frac{1}{m_1+m_2-2} (S_{Y_1} + S_{Y_2}),
\]

where \( S_{Y_j} = \sum_{k=1}^{m_j} (y_{kj} - \bar{y}_j)^2, \ j = 1,2 \).

This point must be stressed that if Model 1 \((j = 1)\) achieves a lower \( MSE \) than the Original Model, and hence if Model 2 \((j = 2)\) achieves an \( MSE \) smaller than that of the Original Model, it would be reasonable to conclude that Model \( 1,2 \) \((j_1 = 1, j_2 = 2)\) will follow suit. The opposite of this argument is not necessarily true. In fact, it is possible for either of Model 1 or Model 2 to have an \( MSE \) higher than that of the Original Model while Model 1,2 achieves a lower \( MSE \) than the Original Model. Under these circumstances, it can be asserted that either of Model 1 or Model 2 will certainly lower the \( MSE \) as that achieved by the Original Model. This point can be explained in mathematical terms as follows.

If \( MSE_0 < S_{1}^2 \Rightarrow (m_1 - 1)MSE_0 < S_{Y_1} \), and

\[
MSE_0 < S_{2}^2 \Rightarrow (m_2 - 1)MSE_0 < S_{Y_2}, \text{ then}
\]

\[
MSE_0 < \frac{1}{m_1+m_2-2} (S_{Y_1} + S_{Y_2}).
\]
The next important point is that if both \( \text{Models 1 and 2} \) achieve lower MSE than the MSE provided by the \textit{Original Model}, it can be concluded that \textit{Model 1,2} achieves an MSE lower than what provided by both \textit{Model 1} and \textit{Model 2}. This is because: if \( \text{MSE}_0 < S_1^2 \Rightarrow \text{MSE}_1 < \text{MSE}_0 \) and since \( \text{MSE}_0 < S_2^2 \), then \( \text{MSE}_1 < \text{MSE}_0 < S_2^2 \). Therefore we obtain

\[ \text{MSE}_{1,2} < \text{MSE}_1. \]

In like manner, if \( \text{MSE}_0 < S_2^2 \Rightarrow \text{MSE}_2 < \text{MSE}_0 \), and since \( \text{MSE}_0 < S_1^2 \) then \( \text{MSE}_2 < \text{MSE}_0 < S_1^2 \). Therefore we obtain

\[ \text{MSE}_{1,2} < \text{MSE}_2. \]

We conclude this section by pointing to the fact that \textit{Model 1,2} can be directly obtained from the \textit{Original Model}. To show this fact, one can simply multiply both sides of (1) by the following \((n - m_1 - m_2 + 2) \times (n - m_1 - m_2 + 2)\) and \((n - m_1 - m_2 + 2) \times n\) matrix combination

\[
\begin{pmatrix}
1/m_1 & 0 & 0 & \cdots & 0 \\
0 & 1/m_2 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
m_1 \\
m_2 \\
\vdots \\
n - m_1 - m_2
\end{pmatrix}
\begin{pmatrix}
1 & \cdots & 1 \\
0 & \cdots & 0 \\
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
1 \\
0 \\
\vdots \\
0
\end{pmatrix}
\]

### 3.2. Employing the Mean Observation Values at All Levels of the Independent Variable

Here we assume that the number of observations of the response variable at each level of the independent variable is strictly larger than 1. As such, our model labeled as \textit{Model 1,2,...,c} will look like

\[
Y_{1,2,\ldots,c} = X_{1,2,\ldots,c} \beta + \epsilon_{1,2,\ldots,c}
\]

Where

\[
Y_{1,2,\ldots,c} = \begin{pmatrix}
\bar{y}_1 \\
\bar{y}_2 \\
\bar{y}_3 \\
\vdots \\
\bar{y}_c
\end{pmatrix}^T
\]

\[
X_{1,2,\ldots,c} = \begin{pmatrix}
1 & 1 & 1 & \cdots & 1 \\
x_1 & x_2 & x_3 & \cdots & x_c
\end{pmatrix}^T
\]

\[
\beta = \begin{pmatrix}
\beta_0 \\
\beta_1
\end{pmatrix}
\]

\[
\epsilon_{1,2,\ldots,c} = \begin{pmatrix}
\bar{\epsilon}_1 \\
\bar{\epsilon}_2 \\
\bar{\epsilon}_3 \\
\vdots \\
\bar{\epsilon}_c
\end{pmatrix}^T
\]

in which \( Y_{1,2,\ldots,c} \) and \( \epsilon_{1,2,\ldots,c} \) are \( c \times 1 \) matrices and \( X_{1,2,\ldots,c} \) is a \( c \times 2 \) matrix. One can arrive at this model by multiplying both sides of (1) by the following \( c \times c \) and \( c \times n \) matrix combination
Again, it can be simply shown that (2) and (4) hold. Also, we employ that weighted least squares to resolve the lack of homogeneity of the variances of the error terms. The following shows that the fitted model is the same as the original fitted model.

\[
\begin{pmatrix}
1/m_1 & 0 & 0 & \ldots & 0 \\
0 & 1/m_2 & 0 & \ldots & 0 \\
0 & 0 & 1/m_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1/m_{c'}
\end{pmatrix}
\begin{pmatrix}
m_1 \\
m_2 \\
m_c
\end{pmatrix}
= 
\begin{pmatrix}
1 & \ldots & 1 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0
\end{pmatrix}
\begin{pmatrix}
1 & \ldots & 1 \\
0 & \ldots & 0 \\
0 & \ldots & 0 \\
0 & \ldots & 0
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
0 \\
\vdots
\end{pmatrix}

The relations that immediately follow, intend to establish this fine point that \( SSE \) in Model \( 1,2,\ldots,c \) is the lowest value among all possible models’ \( SSE \)’s and is equal to the lack of fit sum of squares (SSLF) in the Original Model. In other words, while \( SSE \) in the Original Model assumes its largest value, it assumes its smallest value in Model \( 1,2,\ldots,c \). We know that

\[
SSE_0 = \text{the pure error sum of squares (SSPE)} + \text{SSLF},
\]

or

\[
SSE_0 = \sum_{j=1}^{c} \sum_{k=1}^{m_{j}} (y_{kj} - \bar{y}_j)^2 + \sum_{j=1}^{c} m_{j} \cdot (\bar{y}_j - \bar{y})^2
\]

then

\[
SSE_{1,2,\ldots,c} = \text{SSLF}.
\]

Therefore, we can write

\[
SSE_{M1,2,\ldots,c} \leq \text{Sum of Squared Error} \leq SSE_0
\]

\[
SST_{M1,2,\ldots,c} \leq \text{Total Sum of Squares} \leq SST_0
\]

All this means that \( R^2 \) assumes its largest value in Model \( 1,2,\ldots,c \) and its smallest value in the Original Model. This assertion follows from the fact that \( R^2_0 = \frac{SSR_0}{SST_0} \), \( R^2_{1,2,\ldots,c} = \frac{SSR_{1,2,\ldots,c}}{SST_{1,2,\ldots,c}} \) and (13) is true.

### 4. IDENTIFYING THE MODEL WITH THE SMALLEST MSE

In this section we explain how one can easily identify the model with smallest MSE from among the existing models. Assuming \( m_j \geq 2 \) at \( p \) levels \( p \leq c \), then there are as many as \( 2^p \) models including the Original Model. The following algorithm is proposed to find the model with smallest MSE.

**Algorithm 1**

**Step 1:** Set \( MSE_{new} = MSE_0 \).
Step 2: Compute $S_j^2$ and check the condition $MSE_{new} < S_j^2$, $\forall j$.

If this inequality is not true for any, or if there is no remaining $j$ to be considered, stop and treat $MSE_{new}$ as the smallest $MSE$ and the corresponding model is the model with smallest $MSE$. Otherwise go to Step 3.

Step 3: Consider the level(s) at which the inequality in Step 2 is satisfied; employ the mean of observations instead of the observations at each level and set $MSE_{new}$ equal to the $MSE$ obtained from this model; return to Step 2 for examining the remaining levels.

The discussion presented in section 3.1 provides the motivation for explaining why this algorithm is expected to identify the model with smallest $MSE$.

It is obvious that $MSE$ by itself is not a suitable criterion in comparing the existing models because the model which lowers the $MSE$ faces a decrease in the degrees of freedom as well. In fact, one should judge based on relation (8) in the sense that the model with smallest value of (8) will be the best linear fit on the basis of the power of the tests as well as the accuracy of the prediction and confidence intervals.

5. IDENTIFYING THE MODEL WITH THE LARGEST ACCURACY

Assuming that at $p$ out of $c$ levels of the independent variable, $m_j \geq 2$ holds, we intend to identify the best model on the basis of power of the tests.

We try to minimize

$$t'_{1 - \frac{\alpha}{2}, \text{d.f.}} \cdot \sqrt{MSE}$$

Based on a table of t-distribution, one can prepare a table for tabulating different values of

$$t'_{1 - \frac{\alpha}{2}, \text{d.f.}} = \frac{1}{\sqrt{\text{d.f.}}} \cdot t_{1 - \frac{\alpha}{2}, \text{d.f.}}$$

for various values of $\alpha$ and degrees of freedom. Thus this problem assumes the following form

$$\text{Min} \quad t'_{1 - \frac{\alpha}{2}, \text{d.f.}} \cdot \sqrt{SSE_0 - \sum_{j=1}^{p} a_j S_j}$$

where $a_j$'s are $p$ decision variables assuming value 0 or 1.

$$\text{d.f.} = n - \sum_{j=1}^{p} m_j a_j + \sum_{j=1}^{p} a_j - 2.$$ 

One way to identify the most powerful and the most accurate model would be comparing the objective function for all possible cases and determining the smallest value of (8) for the existing models.

5.1. The Case of Identical Number of Multiple Observations at Selected Levels of the Independent Variable

Suppose we are dealing with a problem in which there are an identical number of multiple observations, $m \geq 2$, at each of $p, 1 \leq p \leq c$ levels of the regressor variable. This obviously is a
special case of the problem as presented in section 3.1. The difference lies in the fact that for this problem one can come up with a simple algorithm to identify the optimal model based on the power of the tests and the accuracy of the confidence intervals.

The interesting point is that based on the nature of the reduction in the degrees of freedom one can easily identify the optimal model. The reduction of degrees of freedom in Model \( j \), \( j = 1, 2, \ldots, p \) amounts to \( m - 1 \). There are as many as \( \binom{p}{1} = p \) of such models. This number for Model \( f_{j_1,j_2} \) stands at \( \binom{p}{2} = \frac{1}{2}p(p + 1) \) and the reduction in the degrees of freedom is as much as \( 2(m - 1) \). In general, when there are \( j \) levels of the independent variable, the reduction of the degrees of freedom amounts to \( j(m - 1) \) and there will be \( \binom{p}{j} \) of such cases. In models with identical degrees of freedom, the term \( t'_{1-d.f.} \) in \( t'_{1-d.f.} \sqrt{SSE} \) remains the same and it suffices to concentrate on \( SSE \) which consists of two parts: \( SSE_0 \) and a quantity which must be subtracted from \( SSE_0 \).

Since \( SSE_0 \) is the same for all models, then among the models with the same degrees of freedom the model in which the largest quantity is subtracted from \( SSE_0 \) corresponds to the objective function with the smallest value. Thus we propose the following algorithm to identify the optimal model.

**Algorithm 2**

**Step 1:** compute \( S_Y \) at each level and sort them in descending order as

\[
S_Y(p) \geq S_Y(p-1) \geq \cdots \geq S_Y(2) \geq S_Y(1).
\]

**Step 2:** compare the objective function for the following \( p + 1 \) models and deliver the optimal model.

\[
\begin{aligned}
\text{Original Model} \\
\text{Model (p)} \\
\text{Model (p),(p-1)} \\
\text{Model (p),(p-1),(p-2)} \\
\vdots \\
\text{Model (p),(p-1),..., (2)} \\
\text{Model (p),(p-1),..., (2),(1)}
\end{aligned}
\]

where, for example, \( \text{Model (p),(p-1)} \) represents the model at levels \( (p) \) and \( (p-1) \) such that \( (p) = \) the level with the largest value of \( S_Y \) and \( (p-1) = \) the level with the second largest value of \( S_Y \).

It must be noted that this algorithm needs just \( p + 1 \) and not \( 2^p \) comparisons to deliver the optimal model.

**6. NUMERICAL EVALUATIONS**

In this section we provide a detailed example to shed light on the merits of our approach. To this end we borrow the following example from Neter et al. (1996).
Example (Problem 20, page 38): **Calculator maintenance.** The Tri-City Office Equipment Corporation sells an imported desk calculator on a franchise basis and performs preventive maintenance and repair service on this calculator. The data below have been collected from 18 recent calls on users to perform routine preventive maintenance service; for each call, $x$ is the number of machines serviced and $y$ is the total number of minutes spent by the service person.

Assume that first-order regression model $(y_i = \beta_0 + \beta_1 x_i + \epsilon_i)$ is appropriate.

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_i$</td>
<td>98</td>
<td>86</td>
<td>78</td>
<td>75</td>
<td>62</td>
<td>101</td>
<td>39</td>
<td>53</td>
<td>33</td>
<td>65</td>
<td>25</td>
<td>71</td>
<td>105</td>
<td>17</td>
<td>49</td>
<td>68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 1 shows, there are 18 pairs of observations in this problem gathered at 8 levels of the regressor variable. At 5 levels of $X$ we have more than one observation on the dependent variable.

Table 1 values of $S_j^2$ for $j = 1, ..., 8$

<table>
<thead>
<tr>
<th>$x_j$</th>
<th>$y_{kj}$</th>
<th>$\hat{y}_j$</th>
<th>$S_{y_j}$</th>
<th>$S_j^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10, 17</td>
<td>13.500</td>
<td>24.500</td>
<td>24.500</td>
</tr>
<tr>
<td>2</td>
<td>33, 25</td>
<td>29.000</td>
<td>32.000</td>
<td>32.000</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>62, 53, 49</td>
<td>54.667</td>
<td>88.667</td>
<td>44.333</td>
</tr>
<tr>
<td>5</td>
<td>78, 75, 65, 71, 68</td>
<td>71.400</td>
<td>109.200</td>
<td>27.300</td>
</tr>
<tr>
<td>6</td>
<td>86</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>101, 98, 105</td>
<td>101.333</td>
<td>24.667</td>
<td>12.333</td>
</tr>
<tr>
<td>8</td>
<td>118</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

By feeding these data to SAS (Statistical Analysis System) we arrive at the following model

$$\hat{y}_j = -2.322 + 14.738x_j$$

Table 2 displays the outcome of all calculations for all possible cases. Figures 4 and 5 display the same results graphically.
The most common values of 1% and 5% for $\alpha$ are used in Table 2, Figure 4, and Figure 5.

Table 2 Summary of calculations for 32 models

<table>
<thead>
<tr>
<th>Model</th>
<th>$SST$</th>
<th>$SSR$</th>
<th>$SSE$</th>
<th>$MSE$</th>
<th>$t_{0.975.d.f. \cdot \sqrt{MSE}}$</th>
<th>$t_{0.995.d.f. \cdot \sqrt{MSE}}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,4,5,7</td>
<td>16224.966</td>
<td>16182.604</td>
<td>42.362</td>
<td>7.060</td>
<td>6.502</td>
<td>9.850</td>
<td>0.997</td>
</tr>
<tr>
<td>1,2,4,5</td>
<td>16249.633</td>
<td>16182.604</td>
<td>67.029</td>
<td>8.379</td>
<td>6.675</td>
<td>9.711</td>
<td>0.996</td>
</tr>
<tr>
<td>2,4,5,7</td>
<td>16249.466</td>
<td>16182.604</td>
<td>91.529</td>
<td>10.170</td>
<td>7.214</td>
<td>10.364</td>
<td>0.994</td>
</tr>
<tr>
<td>2,4,5</td>
<td>16274.133</td>
<td>16182.604</td>
<td>66.862</td>
<td>9.552</td>
<td>7.309</td>
<td>10.814</td>
<td>0.996</td>
</tr>
<tr>
<td>1,4,5</td>
<td>16281.633</td>
<td>16182.604</td>
<td>99.029</td>
<td>11.003</td>
<td>7.503</td>
<td>10.781</td>
<td>0.994</td>
</tr>
<tr>
<td>1,4,5,7</td>
<td>16256.966</td>
<td>16182.604</td>
<td>74.362</td>
<td>10.623</td>
<td>7.708</td>
<td>11.404</td>
<td>0.995</td>
</tr>
<tr>
<td>4,5</td>
<td>16306.133</td>
<td>16182.604</td>
<td>123.529</td>
<td>14.686</td>
<td>8.350</td>
<td>11.707</td>
<td>0.989</td>
</tr>
<tr>
<td>4,5,7</td>
<td>16281.466</td>
<td>16182.604</td>
<td>98.862</td>
<td>12.358</td>
<td>8.106</td>
<td>11.794</td>
<td>0.994</td>
</tr>
<tr>
<td>2,4,5,7</td>
<td>16274.133</td>
<td>16182.604</td>
<td>200.729</td>
<td>15.441</td>
<td>8.488</td>
<td>11.836</td>
<td>0.988</td>
</tr>
<tr>
<td>2,4,5</td>
<td>16306.133</td>
<td>16182.604</td>
<td>208.229</td>
<td>16.018</td>
<td>8.645</td>
<td>12.055</td>
<td>0.987</td>
</tr>
<tr>
<td>1,4</td>
<td>16390.833</td>
<td>16182.604</td>
<td>151.562</td>
<td>15.156</td>
<td>8.674</td>
<td>12.337</td>
<td>0.991</td>
</tr>
<tr>
<td>1,2,4,5,7</td>
<td>16334.166</td>
<td>16182.604</td>
<td>232.729</td>
<td>16.623</td>
<td>8.791</td>
<td>12.504</td>
<td>0.990</td>
</tr>
<tr>
<td>1,2,4,5</td>
<td>16338.300</td>
<td>16182.604</td>
<td>155.696</td>
<td>15.570</td>
<td>8.916</td>
<td>12.426</td>
<td>0.989</td>
</tr>
<tr>
<td>2,4,5</td>
<td>16415.333</td>
<td>16182.604</td>
<td>176.062</td>
<td>16.006</td>
<td>8.806</td>
<td>12.571</td>
<td>0.989</td>
</tr>
<tr>
<td>2,5,7</td>
<td>16438.133</td>
<td>16182.604</td>
<td>180.196</td>
<td>16.381</td>
<td>8.908</td>
<td>12.688</td>
<td>0.989</td>
</tr>
<tr>
<td>1,4,7</td>
<td>16390.833</td>
<td>16182.604</td>
<td>183.562</td>
<td>16.678</td>
<td>8.991</td>
<td>12.721</td>
<td>0.987</td>
</tr>
<tr>
<td>1,5</td>
<td>16370.300</td>
<td>16182.604</td>
<td>208.062</td>
<td>17.338</td>
<td>9.073</td>
<td>12.830</td>
<td>0.989</td>
</tr>
<tr>
<td>1,2,5,7</td>
<td>16313.633</td>
<td>16182.604</td>
<td>187.696</td>
<td>17.063</td>
<td>9.092</td>
<td>12.847</td>
<td>0.987</td>
</tr>
<tr>
<td>1,2,4</td>
<td>16390.666</td>
<td>16182.604</td>
<td>212.196</td>
<td>17.683</td>
<td>9.163</td>
<td>12.949</td>
<td>0.984</td>
</tr>
<tr>
<td>2,5</td>
<td>16454.833</td>
<td>16182.604</td>
<td>264.896</td>
<td>18.921</td>
<td>9.330</td>
<td>13.578</td>
<td>0.992</td>
</tr>
<tr>
<td>1,2,7</td>
<td>16447.500</td>
<td>16182.604</td>
<td>289.396</td>
<td>19.293</td>
<td>9.360</td>
<td>13.723</td>
<td>0.982</td>
</tr>
<tr>
<td>2,7</td>
<td>16447.333</td>
<td>16182.604</td>
<td>289.396</td>
<td>19.293</td>
<td>9.360</td>
<td>13.723</td>
<td>0.982</td>
</tr>
<tr>
<td>1,7</td>
<td>16454.833</td>
<td>16182.604</td>
<td>309.729</td>
<td>19.755</td>
<td>9.502</td>
<td>13.992</td>
<td>0.982</td>
</tr>
<tr>
<td>7</td>
<td>16479.333</td>
<td>16182.604</td>
<td>272.229</td>
<td>20.941</td>
<td>9.884</td>
<td>13.783</td>
<td>0.983</td>
</tr>
</tbody>
</table>

* 0 represents the Original Model

Figure 5 Comparing the relation (8) for all 32 possible cases for $\alpha = 0.05$
As can be seen, for $\alpha = 0.01$ and $\alpha = 0.05$ Model 1,2,4,5 and Model 1,2,4,5,7 are the best models respectively. Also, the ratio of the lengths confidence and prediction intervals of the proposed best models to those of the Original Model which is commonly used in practice are

$$
\alpha = 0.05 : \quad t_{0.975; \text{d.f.}} \cdot \sqrt{\text{MSE}} = \begin{cases} 6.502 & \text{Model 1,2,4,5,7} \\ 9.648 & \text{Original Model} \end{cases} \quad \Rightarrow \frac{6.502}{9.648} = 0.673
$$

$$
\alpha = 0.01 : \quad t_{0.995; \text{d.f.}} \cdot \sqrt{\text{MSE}} = \begin{cases} 9.711 & \text{Model 1,2,4,5} \\ 13.723 & \text{Original Model} \end{cases} \quad \Rightarrow \frac{9.711}{13.723} = 0.708
$$

All this means that Model 1,2,4,5 and Model 1,2,4,5,7 are capable of reducing the lengths of the confidence and prediction intervals by $\%29.2$ and $\%32.7$ for $\alpha = 0.01$ and $\alpha = 0.05$ respectively when compared to the Original Model without any cost.

7. CONCLUSIONS

In this work we showed how one can improve the accuracy of the confidence and prediction intervals in simple linear regression at no cost.

Our treatment has been confined to the problems with multiple observations on the dependent variable at some levels of the regressor variable. In fact, by presenting an algorithm we showed that it suffices to identify the model with smallest value of $t_{1-\frac{\alpha}{2}; \text{d.f.}} \cdot \sqrt{\text{MSE}}$ at a given level of $\alpha$, to arrive at more accurate confidence and prediction intervals.

Extensions to this work consist of designing a more sophisticated algorithm to identify the model with the smallest $t_{1-\frac{\alpha}{2}; \text{d.f.}} \cdot \sqrt{\text{MSE}}$; designing statistical tests and test statistics for comparing different models; and investigating the multiple regression models. Developing a computer code in R system to implement this approach is another avenue for future research.

REFERENCES


